WE CLAIM:

1. A method of detecting and summarising at least one topic in at least one document of a document set, each document in said document set having a plurality of terms and a plurality of sentences comprising said plurality of terms, wherein said plurality of terms and said plurality of sentences are represented as a plurality of vectors in a two-dimensional space, said method comprising the steps of:

pre-processing said at least one document to extract a plurality of significant terms and to create a plurality of basic terms;

formatting said at least one document and said plurality of basic terms;

reducing said plurality of basic terms;

reducing said plurality of sentences;

creating a matrix of said reduced plurality of basic terms and said reduced plurality of sentences;

utilising said matrix to correlate said plurality of basic terms;

transforming a two-dimensional coordinate associated with each of said correlated plurality of basic terms to an n-dimensional coordinate;

clustering said reduced plurality of sentence vectors in said n-dimensional space; and

associating magnitudes of said reduced plurality of sentence vectors with said at least one topic.

2. A method as claimed in Claim 1, wherein said formatting step further comprises producing a file comprising at least one term and an associated location within said at least one document of said at least one term.

3. A method as claimed in Claim 2, wherein said creating step further comprises the steps of:

reading said plurality of basic terms into a term vector;

reading said file comprising at least one term into a document vector;

utilising said term vector, said document vector and an associated threshold to reduce said plurality of basic terms;

utilising said extracted plurality of significant terms to reduce said plurality of sentences; and

reading said reduced plurality of sentences into a sentence vector.

4. A method as claimed in Claim 1, wherein said correlated plurality of basic terms are transformed to hyper spherical coordinates.

5. A method as claimed in Claim 1, wherein end points associated with reduced plurality of sentence vectors lying in close proximity, are clustered.

6. A method as claimed in Claim 5, wherein clusters of said plurality of sentence vectors are linearly shaped.

7. A method as claimed in Claim 6, wherein each of said clusters represents said at least one topic.

8. A method as claimed in Claim 7, wherein field weighting is carried out.

9. A method as claimed in Claim 1, wherein a reduced sentence vector having a large associated magnitude, is associated with at least one topic.

10. A system for detecting and summarising at least one topic in at least one document of a document set, each document in said document set having a plurality of terms and a plurality of sentences comprising said plurality of terms, wherein said plurality of terms and said plurality of sentences are represented as a plurality of vectors in a two-dimensional space, said system comprising:

means for pre-processing said at least one document to extract a plurality of significant terms and to create a plurality of basic terms;

means for formatting said at least one document and said plurality of basic terms;

means for reducing said plurality of basic terms;

means for reducing said plurality of sentences;

means for creating a matrix of said reduced plurality of basic terms on said reduced plurality of sentences;

means for utilising said matrix to correlate said plurality of basic terms;

means for transforming a two-dimensional coordinate associated with each of said correlated plurality of basic terms to an n-dimensional co-ordinate;

means for clustering said reduced plurality of sentence vectors in said n-dimensional space; and

means for associating magnitudes of said reduced plurality of sentence vectors with said at least one topic.

11. Computer readable code stored on a computer readable storage medium for detecting and summarising at least one topic in at least one document of a document set, each document in said document set having a plurality of terms and a plurality of sentences comprising said plurality of terms, said computer readable code comprising:

first processes for pre-processing said at least one document to extract a plurality of significant terms and to create a plurality of basic terms;

second processes for formatting said at least one document and said plurality of basic terms;

third processes for reducing said plurality of basic terms;

fourth processes for reducing said plurality of sentences;

fifth processess for creating a matrix of said reduced plurality of basic terms and said reduced plurality of sentences;

sixth processes for utilising said matrix to correlate said plurality of basic terms;

seventh processes for transforming a two-dimensional coordinate associated with each of said correlated plurality of basic terms to an n-dimensional coordinate;

eighth processess for clustering said reduced plurality of sentence vectors in said n-dimensional space; and

ninth processes associating magnitudes of said reduced plurality of sentence vectors with said at least one topic.

5